

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-304187

(43)Date of publication of application : 18.10.2002

(51)Int.Cl. G10L 13/08  
A63H 11/00  
G10L 13/00  
G10L 15/00

(21)Application number : 2001-104326

(71)Applicant : SONY CORP

(22)Date of filing : 03.04.2001

(72)Inventor : YAMADA KEIICHI  
YAMAZAKI NOBUHIDE  
NITSUTA TOMOAKI  
KOBAYASHI ERIKA  
KOBAYASHI KENICHIRO  
AKAHA MAKOTO

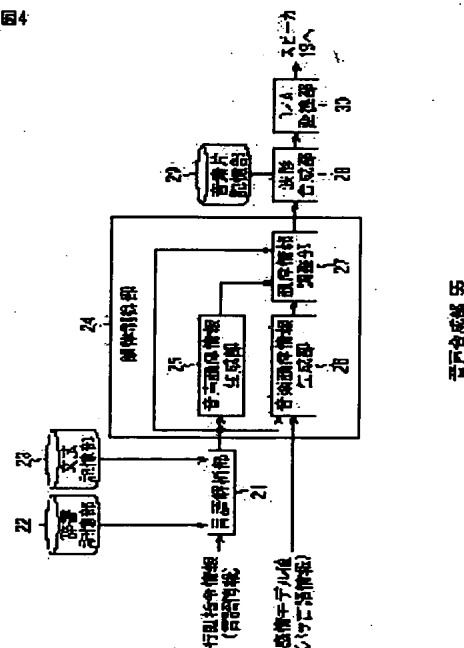
## (54) DEVICE AND METHOD FOR SYNTHESIZING VOICE, PROGRAM AND RECORDING MEDIUM

## (57)Abstract:

PROBLEM TO BE SOLVED: To easily generate synthesized sounds through which paralinguistic information is transmitted effectively.

SOLUTION: A voice rhythm information generating section 25 generates voice rhythm information, that controls the rhythm of synthesized sound based on voice symbols outputted by a language analysis section 21. A music rhythm information generating section 26 controls rhythm by musical elements and generates musical rhythm information to add paralinguistic information, that expresses the feeling of a robot. A rhythm information adjusting section 27 generates final rhythm information, by adjusting the voice rhythm information and the musical rhythm information. A waveform-synthesizing section 28 generates synthesized sounds in which rhythm is controlled by the final rhythm information.

図4



## LEGAL STATUS

-[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision  
of rejection]

[Date of requesting appeal against examiner's  
decision of rejection]

[Date of extinction of right]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号  
特開2002-304187  
(P2002-304187A)

(43) 公開日 平成14年10月18日 (2002.10.18)

(51) Int.Cl. <sup>7</sup>	識別記号	F I	キーワード* (参考)
G 1 0 L 13/08		A 6 3 H 11/00	Z 2 C 1 5 0
A 6 3 H 11/00		C 1 0 L 3/00	H 5 D 0 1 5
G 1 0 L 13/00			Q 5 D 0 4 5
15/00			5 5 1 H

審査請求 未請求 請求項の数11 O L (全 14 頁)

(21) 出願番号	特願2001-104326(P2001-104326)	(71) 出願人	000002185 ソニー株式会社 東京都品川区北品川6丁目7番35号
(22) 出願日	平成13年4月3日(2001.4.3)	(72) 発明者	山田 敬一 東京都品川区北品川6丁目7番35号 ソニー株式会社内
		(72) 発明者	山崎 信英 東京都品川区北品川6丁目7番35号 ソニー株式会社内
		(74) 代理人	100082131 弁理士 稲本 義雄

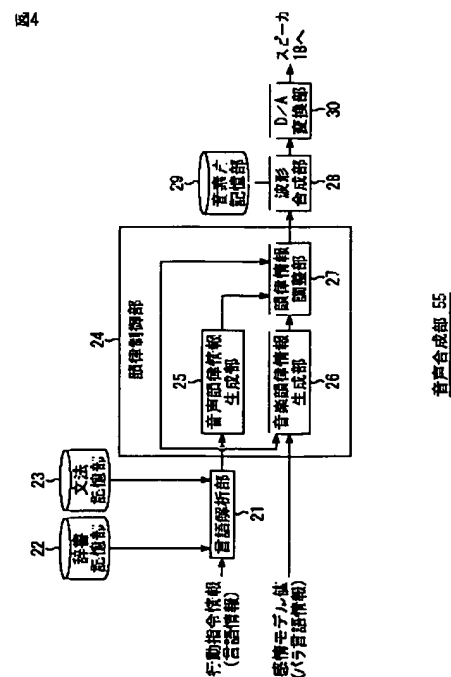
最終頁に続く

(54) 【発明の名称】 音声合成装置および音声合成方法、並びにプログラムおよび記録媒体

(57) 【要約】

【課題】 パラ言語情報を、効果的に伝達することのできる合成音を、容易に生成する。

【解決手段】 音声韻律情報生成部25は、言語解析部21が出力する音声記号に基づいて、合成音の韻律を制御する音声韻律情報を生成し、音楽韻律情報生成部26は、音楽的要素により韻律を制御し、ロボットの感情を表現するパラ言語情報を付与する音楽韻律情報を生成する。韻律情報調整部27は、音声韻律情報と音楽韻律情報を調整することにより、最終的な韻律情報を生成する。そして、波形合成部28は、その最終的な韻律情報により韻律を制御した合成音を生成する。



## 【特許請求の範囲】

【請求項１】 入力データに対応する合成音を生成する音声合成装置であって、  
前記入力データに基づいて、前記合成音の韻律を制御する第１の韻律情報を生成する第１の韻律情報生成手段と、  
音楽的要素により韻律を制御し、前記合成音に対してパラ言語情報を付与する第２の韻律情報を生成する第２の韻律情報生成手段と、  
前記第１と第２の韻律情報を調整し、最終的な韻律情報を生成する韻律情報調整手段と、  
前記最終的な韻律情報により韻律を制御した合成音を生成する音声合成手段とを備えることを特徴とする音声合成装置。

【請求項２】 前記第２の韻律情報生成手段は、旋律、音階、和音、または拍子もしくは律動のうちの１以上の音楽的要素により韻律を制御し、前記合成音に対してパラ言語情報を付与する第２の韻律情報を生成することを特徴とする請求項１に記載の音声合成装置。

【請求項３】 前記第２の韻律情報生成手段は、前記合成音に対して、感情を表現する前記パラ言語情報を付与する第２の韻律情報を生成することを特徴とする請求項１に記載の音声合成装置。

【請求項４】 前記第２の韻律情報生成手段は、前記第２の韻律情報として、MIDI (Musical Instrument Digital Interface) の規格に準拠した情報を生成することを特徴とする請求項１に記載の音声合成装置。

【請求項５】 前記韻律情報調整手段は、前記第２の韻律情報を、前記最終的な韻律情報とすることを特徴とする請求項１に記載の音声合成装置。

【請求項６】 前記韻律情報調整手段は、前記第１の韻律情報におけるピッチに関する情報を、前記第２の韻律情報を利用したものに置き換えることにより、前記最終的な韻律情報を生成することを特徴とする請求項１に記載の音声合成装置。

【請求項７】 前記韻律情報調整手段は、前記第１の韻律情報の一部を、対応する前記第２の韻律情報の一部と置き換えることにより、前記最終的な韻律情報を生成することを特徴とする請求項１に記載の音声合成装置。

【請求項８】 前記韻律情報調整手段は、前記第１の韻律情報における音韻継続時間長を、前記第２の韻律情報を利用したものに置き換えることにより、前記最終的な韻律情報を生成することを特徴とする請求項１に記載の音声合成装置。

【請求項９】 入力データに対応する合成音を生成する音声合成方法であって、  
前記入力データに基づいて、前記合成音の韻律を制御する第１の韻律情報を生成する第１の韻律情報生成ステップと、  
音楽的要素により韻律を制御し、前記合成音に対してパ

ラ言語情報を付与する第２の韻律情報を生成する第２の韻律情報生成ステップと、

前記第１と第２の韻律情報を調整し、最終的な韻律情報を生成する韻律情報調整ステップと、  
前記最終的な韻律情報により韻律を制御した合成音を生成する音声合成ステップとを備えることを特徴とする音声合成方法。

【請求項１０】 入力データに対応する合成音を生成する音声合成処理を、コンピュータに行わせるプログラムであって、  
前記入力データに基づいて、前記合成音の韻律を制御する第１の韻律情報を生成する第１の韻律情報生成ステップと、

音楽的要素により韻律を制御し、前記合成音に対してパラ言語情報を付与する第２の韻律情報を生成する第２の韻律情報生成ステップと、

前記第１と第２の韻律情報を調整し、最終的な韻律情報を生成する韻律情報調整ステップと、  
前記最終的な韻律情報により韻律を制御した合成音を生成する音声合成ステップとを備えることを特徴とするプログラム。

【請求項１１】 入力データに対応する合成音を生成する音声合成処理を、コンピュータに行わせるプログラムが記録されている記録媒体であって、

前記入力データに基づいて、前記合成音の韻律を制御する第１の韻律情報を生成する第１の韻律情報生成ステップと、

音楽的要素により韻律を制御し、前記合成音に対してパラ言語情報を付与する第２の韻律情報を生成する第２の韻律情報生成ステップと、

前記第１と第２の韻律情報を調整し、最終的な韻律情報を生成する韻律情報調整ステップと、  
前記最終的な韻律情報により韻律を制御した合成音を生成する音声合成ステップとを備えるプログラムが記録されていることを特徴とする記録媒体。

## 【発明の詳細な説明】

【０００１】

【発明の属する技術分野】 本発明は、音声合成装置および音声合成方法、並びにプログラムおよび記録媒体に関する。特に、例えば、パラ言語情報を、効果的に伝達することのできる合成音を、容易に生成することができるようにする音声合成装置および音声合成方法、並びにプログラムおよび記録媒体に関する。

【０００２】

【従来の技術】 人間の発話である自然音声には、言語上の意味を伝達する言語情報の他に、その発話者の個性（個人性）や、意志（意図）、感情等を表現するパラ言語情報が含まれる。

【０００３】 一方、音声合成装置では、例えば、テキストから合成音を生成する音声合成処理（ＴＴＳ (Text-To

-Speech)処理)の際に、自然音声により近い合成音を生成するために、合成音に対して、パラ言語情報が付与される。

【0004】即ち、音声合成装置では、合成音におけるピッチ周波数の時間変化であるピッチパターンや、合成音のパワーの時間変化であるパワーパターン、合成音を構成する各音韻の継続時間長(音韻継続時間長)といった韻律を制御することにより、感情等を表現するパラ言語情報が、合成音に付与される。

【0005】なお、ピッチパターンによるピッチ周波数により、合成音による声の高さが規定される。また、パワーパターンによるパワーにより、合成音による声の強弱が規定される。さらに、音韻継続時間長により、合成音の発話速度やリズム等が規定される。

【0006】

【発明が解決しようとする課題】合成音に対して、自然音声に含まれるのと同様のパラ言語情報を付与することができれば、自然音声に近い合成音を得ることが可能となる。

【0007】しかしながら、例えば、規則音声合成のように、ある規則にしたがって音声合成を行う場合等においては、自然音声に含まれるのと同様のパラ言語情報を付与する規則を記述することは容易ではなく、従って、パラ言語情報を、効果的に伝達することのできる合成音を、容易に生成することのできる手法の提案が要請されている。

【0008】本発明は、このような状況に鑑みてなされたものであり、パラ言語情報を、効果的に伝達することのできる合成音を、容易に生成することができるようにするものである。

【0009】

【課題を解決するための手段】本発明の音声合成装置は、入力データに基づいて、合成音の韻律を制御する第1の韻律情報を生成する第1の韻律情報生成手段と、音楽的要素により韻律を制御し、合成音に対してパラ言語情報を付与する第2の韻律情報を生成する第2の韻律情報生成手段と、第1と第2の韻律情報を調整し、最終的な韻律情報を生成する韻律情報調整手段と、最終的な韻律情報により韻律を制御した合成音を生成する音声合成手段とを備えることを特徴とする。

【0010】本発明の音声合成方法は、入力データに基づいて、合成音の韻律を制御する第1の韻律情報を生成する第1の韻律情報生成ステップと、音楽的要素により韻律を制御し、合成音に対してパラ言語情報を付与する第2の韻律情報を生成する第2の韻律情報生成ステップと、第1と第2の韻律情報を調整し、最終的な韻律情報を生成する韻律情報調整ステップと、最終的な韻律情報により韻律を制御した合成音を生成する音声合成ステップとを備えることを特徴とする。

【0011】本発明のプログラムは、入力データに基づ

いて、合成音の韻律を制御する第1の韻律情報を生成する第1の韻律情報生成ステップと、音楽的要素により韻律を制御し、合成音に対してパラ言語情報を付与する第2の韻律情報を生成する第2の韻律情報生成ステップと、第1と第2の韻律情報を調整し、最終的な韻律情報を生成する韻律情報調整ステップと、最終的な韻律情報により韻律を制御した合成音を生成する音声合成ステップとを備えることを特徴とする。

【0012】本発明の記録媒体は、入力データに基づいて、合成音の韻律を制御する第1の韻律情報を生成する第1の韻律情報生成ステップと、音楽的要素により韻律を制御し、合成音に対してパラ言語情報を付与する第2の韻律情報を生成する第2の韻律情報生成ステップと、第1と第2の韻律情報を調整し、最終的な韻律情報を生成する韻律情報調整ステップと、最終的な韻律情報により韻律を制御した合成音を生成する音声合成ステップとを備えるプログラムが記録されていることを特徴とする。

【0013】本発明の音声合成装置および音声合成方法、並びにプログラムにおいては、入力データに基づいて、合成音の韻律を制御する第1の韻律情報が生成されるとともに、音楽的要素により韻律を制御し、合成音に対してパラ言語情報を付与する第2の韻律情報が生成される。さらに、第1と第2の韻律情報が調整され、最終的な韻律情報が生成される。そして、その最終的な韻律情報により韻律を制御した合成音が生成される。

【0014】

【発明の実施の形態】図1は、本発明を適用したロボットの一実施の形態の外観構成例を示しており、図2は、その電気的構成例を示している。

【0015】本実施の形態では、ロボットは、例えば、犬等の四つ足の動物の形状のものとなっており、胴体部ユニット2の前後左右に、それぞれ脚部ユニット3A、3B、3C、3Dが連結されるとともに、胴体部ユニット2の前端部と後端部に、それぞれ頭部ユニット4と尻尾部ユニット5が連結されることにより構成されている。

【0016】尻尾部ユニット5は、胴体部ユニット2の上面に設けられたベース部5Bから、2自由度をもって湾曲または揺動自在に引き出されている。

【0017】胴体部ユニット2には、ロボット全体の制御を行うコントローラ10、ロボットの動力源となるバッテリー11、並びにバッテリーセンサ12および熱センサ13からなる内部センサ部14などが収納されている。

【0018】頭部ユニット4には、「耳」に相当するマイク(マイクロフォン)15、「目」に相当するCCD(Charge Coupled Device)カメラ16、触覚に相当するタッチセンサ17、「口」に相当するスピーカ18などが、それぞれ所定位置に配設されている。また、頭部ユニット4には、口の下顎に相当する下顎部4Aが1自由

度をもって可動に取り付けられており、この下顎部4Aが動くことにより、ロボットの口の開閉動作が実現されるようになっている。

【0019】脚部ユニット3A乃至3Dそれぞれの関節部分や、脚部ユニット3A乃至3Dそれぞれと胴体部ユニット2の連結部分、頭部ユニット4と胴体部ユニット2の連結部分、頭部ユニット4と下顎部4Aの連結部分、並びに尻尾部ユニット5と胴体部ユニット2の連結部分などには、図2に示すように、それぞれアクチュエータ3AA<sub>1</sub>乃至3AA<sub>K</sub>、3BA<sub>1</sub>乃至3BA<sub>K</sub>、3CA<sub>1</sub>乃至3CA<sub>K</sub>、3DA<sub>1</sub>乃至3DA<sub>K</sub>、4A<sub>1</sub>乃至4A<sub>L</sub>、5A<sub>1</sub>および5A<sub>2</sub>が配設されている。

【0020】頭部ユニット4におけるマイク15は、ユーザからの発話を含む周囲の音声(音)を集音し、得られた音声信号を、コントローラ10に送出する。CCDカメラ16は、周囲の状況を撮像し、得られた画像信号を、コントローラ10に送出する。

【0021】タッチセンサ17は、例えば、頭部ユニット4の上部に設けられており、ユーザからの「なでる」や「たたく」といった物理的な働きかけにより受けた圧力を検出し、その検出結果を圧力検出信号としてコントローラ10に送出する。

【0022】胴体部ユニット2におけるバッテリーセンサ12は、バッテリー11の残量を検出し、その検出結果を、バッテリー残量検出信号としてコントローラ10に送出する。熱センサ13は、ロボット内部の熱を検出し、その検出結果を、熱検出信号としてコントローラ10に送出する。

【0023】コントローラ10は、CPU(Central Processing Unit)10Aやメモリ10B等を内蔵しており、CPU10Aにおいて、メモリ10Bに記憶された制御プログラムが実行されることにより、各種の処理を行う。

【0024】即ち、コントローラ10は、マイク15や、CCDカメラ16、タッチセンサ17、バッテリーセンサ12、熱センサ13から与えられる音声信号、画像信号、圧力検出信号、バッテリー残量検出信号、熱検出信号に基づいて、周囲の状況や、ユーザからの指令、ユーザからの働きかけなどの有無を判断する。

【0025】さらに、コントローラ10は、この判断結果等に基づいて、続く行動を決定し、その決定結果に基づいて、アクチュエータ3AA<sub>1</sub>乃至3AA<sub>K</sub>、3BA<sub>1</sub>乃至3BA<sub>K</sub>、3CA<sub>1</sub>乃至3CA<sub>K</sub>、3DA<sub>1</sub>乃至3DA<sub>K</sub>、4A<sub>1</sub>乃至4A<sub>L</sub>、5A<sub>1</sub>、5A<sub>2</sub>のうちの必要ものを駆動させる。これにより、頭部ユニット4を上下左右に振らせたり、下顎部4Aを開閉させる。さらには、尻尾部ユニット5を動かしたり、各脚部ユニット3A乃至3Dを駆動して、ロボットを歩行させるなどの行動を行わせる。

【0026】また、コントローラ10は、必要に応じ

て、合成音を生成し、スピーカ18に供給して出力させたり、ロボットの「目」の位置に設けられた図示しないLED(Light Emitting Diode)を点灯、消灯または点滅させる。

【0027】以上のようにして、ロボットは、周囲の状況等に基づいて自律的に行動をとるようになっている。

【0028】なお、メモリ10Bは、例えば、メモリスティック(商標)等の、容易に着脱が可能なメモリカードで構成することが可能である。

【0029】次に、図3は、図2のコントローラ10の機能的構成例を示している。なお、図3に示す機能的構成は、CPU10Aが、メモリ10Bに記憶された制御プログラムを実行することで実現されるようになっている。

【0030】コントローラ10は、特定の外部状態を認識するセンサ入力処理部50、センサ入力処理部50の認識結果を累積して、感情や、本能、成長の状態を表現するモデル記憶部51、センサ入力処理部50の認識結果等に基づいて、続く行動を決定する行動決定機構部52、行動決定機構部52の決定結果に基づいて、実際にロボットに行動を起こさせる姿勢遷移機構部53、各アクチュエータ3AA<sub>1</sub>乃至5A<sub>1</sub>および5A<sub>2</sub>を駆動制御する制御機構部54、並びに合成音を生成する音声合成部55から構成されている。

【0031】センサ入力処理部50は、マイク15や、CCDカメラ16、タッチセンサ17等から与えられる音声信号、画像信号、圧力検出信号等に基づいて、特定の外部状態や、ユーザからの特定の働きかけ、ユーザからの指示等を認識し、その認識結果を表す状態認識情報を、モデル記憶部51および行動決定機構部52に通知する。

【0032】即ち、センサ入力処理部50は、音声認識部50Aを有しており、音声認識部50Aは、マイク15から与えられる音声信号について音声認識を行う。そして、音声認識部50Aは、その音声認識結果としての、例えば、「歩け」、「伏せ」、「ボールを追いかける」等の指令その他を、状態認識情報として、モデル記憶部51および行動決定機構部52に通知する。

【0033】また、センサ入力処理部50は、画像認識部50Bを有しており、画像認識部50Bは、CCDカメラ16から与えられる画像信号を用いて、画像認識処理を行う。そして、画像認識部50Bは、その処理の結果、例えば、「赤い丸いもの」や、「地面に対して垂直なかつ所定高さ以上の平面」等を検出したときには、「ボールがある」や、「壁がある」等の画像認識結果を、状態認識情報として、モデル記憶部51および行動決定機構部52に通知する。

【0034】さらに、センサ入力処理部50は、圧力処理部50Cを有しており、圧力処理部50Cは、タッチセンサ17から与えられる圧力検出信号を処理する。そ

して、圧力処理部５０Ｃは、その処理の結果、所定の閾値以上で、かつ短時間の圧力を検出したときには、「たたかれた（しかられた）」と認識し、所定の閾値未満で、かつ長時間の圧力を検出したときには、「なでられた（ほめられた）」と認識して、その認識結果を、状態認識情報として、モデル記憶部５１および行動決定機構部５２に通知する。

【0035】モデル記憶部51は、ロボットの感情、本能、成長の状態を表現する感情モデル、本能モデル、成長モデルをそれぞれ記憶、管理している。

【0036】ここで、感情モデルは、例えば、「喜び」、「悲しさ」、「怒り」、「驚き」、「恐れ」、「落胆」、「平静」等の感情の状態(度合い)を、所定の範囲(例えば、-1.0乃至1.0等)の値によってそれぞれ表し、センサ入力処理部50からの状態認識情報や時間経過等に基づいて、その値を変化させる。本能モデルは、例えば、「食欲」、「睡眠欲」、「運動欲」等の本能による欲求の状態(度合い)を、所定の範囲の値によってそれぞれ表し、センサ入力処理部50からの状態認識情報や時間経過等に基づいて、その値を変化させる。成長モデルは、例えば、「幼年期」、「青年期」、「熟年期」、「老年期」等の成長の状態(度合い)を、所定の範囲の値によってそれぞれ表し、センサ入力処理部50からの状態認識情報や時間経過等に基づいて、その値を変化させる。

【0037】モデル記憶部51は、上述のようにして感情モデル、本能モデル、成長モデルの値で表される感情、本能、成長の状態を、状態情報として、行動決定機構部52に送出する。

【0038】なお、モデル記憶部51には、センサ入力処理部50から状態認識情報が供給される他、行動決定機構部52から、ロボットの現在または過去の行動、具体的には、例えば、「長時間歩いた」などの行動の内容を示す行動情報が供給されるようになっており、モデル記憶部51は、同一の状態認識情報が与えられても、行動情報が示すロボットの行動に応じて、異なる状態情報を生成するようになっている。

【0039】即ち、例えば、ロボットが、ユーザに挨拶をし、ユーザに頭を撫でられた場合には、ユーザに挨拶をしたという行動情報と、頭を撫でられたという状態認識情報とが、モデル記憶部51に与えられ、この場合、モデル記憶部51では、「うれしさ」を表す感情モデルの値が増加される。

【0040】一方、ロボットが、何らかの仕事を実行中に頭を撫でられた場合には、仕事を実行中であるという行動情報と、頭を撫でられたという状態認識情報とが、モデル記憶部51に与えられ、この場合、モデル記憶部51では、「うれしさ」を表す感情モデルの値は変化されない。

【0041】このように、モデル記憶部51は、状態認

識情報だけでなく、現在または過去のロボットの行動を示す行動情報も参照しながら、感情モデルの値を設定する。これにより、例えば、何らかのタスクを実行中に、ユーザが、いたずらするつもりで頭を撫でたときに、「うれしさ」を表す感情モデルの値を増加させるような、不自然な感情の変化が生じることを回避することができる。

【0042】なお、モデル記憶部51は、本能モデルおよび成長モデルについても、感情モデルにおける場合と同様に、状態認識情報および行動情報の両方に基づいて、その値を増減させるようになっている。また、モデル記憶部51は、感情モデル、本能モデル、成長モデルそれぞれの値を、他のモデルの値にも基づいて増減させるようになっている。

【0043】行動決定機構部52は、センサ入力処理部50からの状態認識情報や、モデル記憶部51からの状態情報、時間経過等に基づいて、次の行動を決定し、決定された行動の内容を、行動指令情報として、姿勢遷移機構部53に送出する。

【0044】即ち、行動決定機構部52は、ロボットがとり得る行動をステート（状態）(state)に対応させた有限オートマトンを、ロボットの行動を規定する行動モデルとして管理しており、この行動モデルとしての有限オートマトンにおけるステートを、センサ入力処理部50からの状態認識情報や、モデル記憶部51における感情モデル、本能モデル、または成長モデルの値、時間経過等に基づいて遷移させ、遷移後のステートに対応する行動を、次にとるべき行動として決定する。

【0045】ここで、行動決定機構部52は、所定のトリガ(trigger)があったことを検出すると、ステートを遷移させる。即ち、行動決定機構部52は、例えば、現在のステートに対応する行動を実行している時間が所定時間に達したときや、特定の状態認識情報を受信したとき、モデル記憶部51から供給される状態情報が示す感情や、本能、成長の状態の値が所定の閾値以下または以上になったとき等に、ステートを遷移させる。

【0046】なお、行動決定機構部52は、上述したように、センサ入力処理部50からの状態認識情報だけでなく、モデル記憶部51における感情モデルや、本能モデル、成長モデルの値等にも基づいて、行動モデルにおけるステートを遷移させることから、同一の状態認識情報が入力されても、感情モデルや、本能モデル、成長モデルの値（状態情報）によっては、ステートの遷移先は異なるものとなる。

【0047】その結果、行動決定機構部52は、例えば、状態情報が、「怒っていない」こと、および「お腹がすいていない」ことを表している場合において、状態認識情報が、「目の前に手のひらが差し出された」ことを表しているときには、目の前に手のひらが差し出されたことに応じて、「お手」という行動をとらせる行動指

令情報を生成し、これを、姿勢遷移機構部53に送出する。

【0048】また、行動決定機構部52は、例えば、状態情報が、「怒っていない」こと、および「お腹がすいている」ことを表している場合において、状態認識情報が、「目の前に手のひらが差し出された」ことを表しているときには、目の前に手のひらが差し出されたことに応じて、「手のひらをべろべろなめる」ような行動を行わせるための行動指令情報を生成し、これを、姿勢遷移機構部53に送出する。

【0049】また、行動決定機構部52は、例えば、状態情報が、「怒っている」ことを表している場合において、状態認識情報が、「目の前に手のひらが差し出された」ことを表しているときには、状態情報が、「お腹がすいている」ことを表している、また、「お腹がすいていない」ことを表している、また、「ふいと横を向く」ような行動を行わせるための行動指令情報を生成し、これを、姿勢遷移機構部53に送出する。

【0050】なお、行動決定機構部52では、上述したように、ロボットの頭部や手足等を動作させる行動指令情報の他、ロボットに発話を行わせる行動指令情報も生成される。ロボットに発話を行わせる行動指令情報は、音声合成部55に供給されるようになっており、音声合成部55に供給される行動指令情報には、音声合成部55に生成させる合成音に対応する漢字仮名混じり等のテキスト、あるいは音韻情報（読み）と、アクセントやポーズの位置を示すタグからなる音声記号等の言語情報が含まれる。そして、音声合成部55は、行動決定部52から行動指令情報を受信すると、その行動指令情報に含まれる言語情報に基づき、合成音を生成し、スピーカ18に供給して出力させる。これにより、スピーカ18からは、例えば、ロボットの鳴き声、さらには、「お腹がすいた」等のユーザへの各種の要求、「何？」等のユーザの呼びかけに対する応答その他の音声出力が行われる。

【0051】ここで、音声合成部55には、モデル記憶部51から状態情報も供給されるようになっており、音声合成部55は、この状態情報における感情モデルの値（感情モデル値）に基づいて韻律を制御した合成音を生成することが可能となっている。

【0052】なお、音声合成部55では、感情の他、本能や本能の状態に基づいて韻律を制御した合成音を生成することも可能である。また、行動決定機構部52は、合成音を出力する場合には、下顎部4Aを開閉させる行動指令情報を、必要に応じて生成し、姿勢遷移機構部53に出力する。この場合、合成音の出力に同期して、下顎部4Aが開閉し、ユーザに、ロボットがしゃべっているかのような印象を与えることができる。

【0053】姿勢遷移機構部53は、行動決定機構部52から供給される行動指令情報に基づいて、ロボットの

姿勢を、現在の姿勢から次の姿勢に遷移させるための姿勢遷移情報を生成し、これを制御機構部54に送出する。

【0054】ここで、現在の姿勢から次に遷移可能な姿勢は、例えば、胴体や手や足の形状、重さ、各部の結合状態のようなロボットの物理的形状と、関節が曲がる方向や角度のようなアクチュエータ3A<sub>1</sub>乃至5A<sub>1</sub>および5A<sub>2</sub>の機構とによって決定される。

【0055】制御機構部54は、姿勢遷移機構部53からの姿勢遷移情報にしたがって、アクチュエータ3A<sub>1</sub>乃至5A<sub>1</sub>および5A<sub>2</sub>を駆動するための制御信号を生成し、これを、アクチュエータ3A<sub>1</sub>乃至5A<sub>1</sub>および5A<sub>2</sub>に送出する。これにより、アクチュエータ3A<sub>1</sub>乃至5A<sub>1</sub>および5A<sub>2</sub>は、制御信号にしたがって駆動し、ロボットは、自律的に行動を起こす。

【0056】次に、図4は、図3の音声合成部55の構成例を示している。

【0057】言語解析部21には、行動決定機構部52が出力する、音声合成の対象とする言語情報を含む行動指令情報が供給されるようになっており、言語解析部21は、辞書記憶部22や文法記憶部23を参照しながら、その行動指令情報に含まれる言語情報を解析する。

【0058】即ち、辞書記憶部22には、各単語の品詞情報や、読み、アクセント等の情報が記述された単語辞書が記憶されており、また、文法記憶部23には、辞書記憶部22の単語辞書に記述された単語について、単語連鎖に関する制約等の文法規則が記憶されている。そして、言語解析部21は、この単語辞書および文法規則に基づいて、そこに入力される言語情報の形態素解析や構文解析等の言語解析を行うことにより、後段の韻律制御部24および波形合成部28で行われる規則音声合成に必要な情報を抽出する。ここで、規則音声合成に必要な情報としては、例えば、ポーズの位置や、アクセントおよびイントネーションを表す情報や、各単語の発音等の音韻情報などがある。言語解析部21は、言語解析が終了すると、その言語解析の結果得られた情報から、音声記号を生成して出力する。

【0059】なお、言語解析部21は、言語情報がテキストである場合に、上述のような言語解析を行う。即ち、言語情報が音声記号である場合には、一般に、音声記号には、音韻情報と、ポーズの位置やアクセントを表す情報（タグ）とが含まれているため、言語解析部21は、そのような音声記号については、特に処理を行わず、そのまま出力する。

【0060】言語解析部21が出力する音声記号は、韻律制御部24に供給される。韻律制御部24には、さらに、モデル記憶部51（図3）に記憶された感情モデル値が、パラ言語情報として供給されるようになっており、韻律制御部24は、言語解析部21からの音声記号に基づいて、韻律情報を生成するとともに、音楽的要素



により韻律を制御し、合成音に対して、モデル記憶部51からの感情モデル値が表す感情を表現する韻律情報を生成する。さらに、韻律制御部24は、その2つの韻律情報から、最終的に合成音の韻律制御に用いる韻律情報を生成する。

【0061】即ち、韻律制御部24は、音声韻律情報生成部25、音楽韻律情報生成部26、および韻律情報調整部27から構成されている。

【0062】音声韻律情報生成部25は、言語解析部21が出力する音声記号に基づいて、例えば、従来の音声合成装置と同様に、合成音におけるピッチ周波数の時間変化であるピッチパターンや、合成音のパワーの時間変化であるパワーパターン、合成音を構成する各音韻の継続時間長（音韻継続時間長）等の韻律情報を生成し、韻律情報調整部27に供給する。ここで、音声韻律情報生成部25が生成する韻律情報を、以下、適宜、音声韻律情報という。

【0063】音楽韻律情報生成部26は、言語解析部21が出力する音声記号から、合成音の音韻や、ポーズの位置、アクセントの位置等の言語的な情報を認識し、それらの情報と、モデル記憶部51からの感情モデル値に基づいて、旋律（メロディ）、音階（スケール）、和音（ハーモニ）、または拍子（ビート）（もしくは律動（リズム））等の音楽的要素による韻律情報を生成する。

【0064】即ち、音楽韻律情報生成部26は、言語解析部21が出力する音声記号から認識される音韻について、感情モデル値が表す感情とその状態に対応する所定の旋律、音階、和音、または拍子（もしくは律動）のうちの1以上を与えるピッチパターンや、パワーパターン、音韻継続時間長等の韻律情報を生成し、韻律情報調整部27に供給する。

【0065】具体的には、例えば、感情モデル値が、「喜び」の感情が大きいことを表している場合には、音楽韻律情報生成部26は、長調の和音（例えば、「ドミソ」など）や、速い拍子等を与えるピッチパターン、パワーパターン、音韻継続時間長等の韻律情報を生成する。さらに、「喜び」の感情が、より大きい場合には、音楽韻律情報生成部26は、より高い音による和音や、より速い拍子等を与える韻律情報を生成する。

【0066】また、例えば、感情モデル値が、「悲しみ」の感情が大きいことを表している場合には、音楽韻律情報生成部26は、短調の和音（例えば、ミを半音下げた「ド♭ミソ」など）や、遅い拍子等を与えるピッチパターン、パワーパターン、音韻継続時間長等の韻律情報を生成する。

【0067】ここで、音楽韻律情報生成部26が生成する韻律情報を、音声韻律情報生成部25が生成する韻律情報（音声韻律情報）と区別するために、以下、適宜、音楽韻律情報という。

【0068】韻律情報調整部27は、音声韻律情報生成部25からの音声韻律情報と、音楽韻律情報生成部26からの音楽韻律情報を調整し、最終的に、合成音の韻律を制御する韻律情報（以下、適宜、最終韻律情報という）を生成する。

【0069】即ち、韻律情報調整部27は、例えば、音楽韻律情報が、ある特定の旋律を与えるものである場合には、音楽韻律情報を、そのまま最終韻律情報とする。

【0070】従って、この場合、音声韻律情報と音楽韻律情報が、例えば、図5（A）と図5（B）にそれぞれ示すようなものであるときには、韻律情報調整部27は、図5（C）に示すように、図5（B）に示した音楽韻律情報を、そのまま、最終韻律情報とする。

【0071】ここで、図5においては（後述する図6乃至図8においても同様）、テキスト「楽しいな」に対する韻律情報としてのピッチパターン、パワーパターン、および音韻継続時間長を示してある。即ち、図5において、太線で示す線（直線または曲線）は、ピッチパターンを表す。また、アルファベットで示してある各音韻（音素）上に配置してある長方形の縦方向の長さ（高さ）が、パワーを表し、その横方向の長さが、音韻継続時間長を表す。

【0072】その他、韻律情報調整部27は、例えば、音楽韻律情報が、ある特定の音階を与えるものである場合には、音声韻律情報におけるピッチパターンを、音楽韻律情報におけるピッチパターンを利用したもの置き換え、そのピッチパターンの置き換え後の音声韻律情報を、最終韻律情報とする。

【0073】従って、この場合、音声韻律情報と音楽韻律情報が、例えば、図6（A）と図6（B）にそれぞれ示すようなものであるときには、韻律情報調整部27は、図6（C）に示すような韻律情報を、最終韻律情報とする。

【0074】ここで、図6（C）の最終韻律情報は、図6（A）の音声韻律情報における各音韻（音節）のピッチ周波数を、そのピッチ周波数に対し、図6（B）の音楽韻律情報のピッチ周波数の中で最も近いものに置き換えたものとなっている。

【0075】また、韻律情報調整部27は、例えば、音声韻律情報の一部を、その一部に対応する音楽韻律情報の一部と置き換え、その置き換え後の音声韻律情報を、最終韻律情報とする。

【0076】従って、この場合、音声韻律情報と音楽韻律情報が、例えば、図7（A）と図7（B）にそれぞれ示すようなものであるときには、韻律情報調整部27は、図7（C）に示すような韻律情報を、最終韻律情報とする。

【0077】ここで、図7（C）の最終韻律情報は、図7（A）の音声韻律情報における最後の音節（na）の部分を、図7（B）の音楽韻律情報における最後の音節の

部分に置き換えたものとなっている。

【0078】さらに、韻律情報調整部27は、例えば、音楽韻律情報が、ある特定の律動（あるいは拍子）を与えるものである場合には、音声韻律情報における音韻継続時間長を、音楽韻律情報における音韻継続時間長に置き換え、その音韻継続時間長の置き換え後の音声韻律情報を、最終韻律情報とする。

【0079】従って、この場合、音声韻律情報と音楽韻律情報が、例えば、図8（A）と図8（B）にそれぞれ示すようなものであるときには、韻律情報調整部27は、図8（C）に示すような韻律情報を、最終韻律情報とする。

【0080】ここで、図8（C）の最終韻律情報は、図8（A）の音声韻律情報における各音韻の音韻継続時間長を、図8（B）の音楽韻律情報における各音韻の音韻継続時間長に置き換えたものとなっている。さらに、音韻継続時間長の置き換えにより、合成音の発話時間（音声区間）が変化することから、図8（C）の最終韻律情報は、そのような発話時間の変化に伴い、図8（A）の音声韻律情報におけるピッチパターンが時間軸伸縮されたものとなっている。

【0081】韻律情報調整部27は、上述のようにして、音声韻律情報と音楽韻律情報を調整することにより、最終韻律情報を生成し、言語解析部21から供給される音声記号とともに、波形合成部28に出力する。

【0082】なお、音声韻律情報と音楽韻律情報を調整して最終韻律情報を生成する方法は、上述した方法に限定されるものではない。

【0083】波形合成部28は、音素片記憶部29を参照し、韻律情報調整部27からの最終韻律情報により韻律を制御した、同じく韻律情報調整部27からの音声記号における音韻情報に対応する合成音データ（デジタルデータ）を生成する。

【0084】即ち、音素片記憶部29には、例えば、C V (Consonant, Vowel) や、V C V、C V C等の形で音素片データが記憶されており、波形合成部28は、韻律情報調整部27からの音声記号における音韻情報に基づいて、必要な音素片データを接続し、さらに、ポーズ、アクセント、イントネーション等を、韻律情報調整部27からの最終韻律情報にしたがって付加することで、言語解析部21に入力された言語情報に対応する合成音データを生成する。

【0085】この合成音データは、D/A (Digital/Analog) 変換部30に供給される。D/A変換部30は、デジタル信号としての合成音データを、アナログ信号としての音声信号にD/A変換し、スピーカ18に供給する。これにより、言語解析部21に入力された言語情報に対応し、かつ、音楽的要素によるパラ言語情報であって、ロボットの感情を表現するものが付与された合成音出力される。

【0086】次に、図9のフローチャートを参照して、図4の音声合成部55による音声合成処理について説明する。

【0087】行動決定機構部52（図3）が、音声合成の対象とする言語情報を含む行動指令情報を、音声合成部55に出力すると、言語解析部21は、その行動指令情報を受信する。そして、言語解析部21は、その行動指令情報に含まれる言語情報について、言語解析を行い、その結果得られる音声記号を、韻律制御部24の音声韻律情報生成部25、音楽韻律情報生成部26、および韻律情報調整部27に供給する。

【0088】その後、ステップS2に進み、音声韻律情報生成部25と音楽韻律情報生成部26は、韻律情報を生成する。

【0089】即ち、音声韻律情報生成部25は、言語解析部21が出力する音声記号に基づいて、例えば、標準的な音声韻律情報を生成し、韻律情報調整部27に供給する。

【0090】また、音楽韻律情報生成部26は、モデル記憶部51（図3）から感情モデル値を、パラ言語情報として読み出す。さらに、音楽韻律情報生成部26は、言語解析部21が出力する音声記号から認識される音韻について、感情モデル値が表す感情とその状態に対応する所定の旋律、音階、和音、または拍子（もしくは律動）を与える音楽韻律情報を生成し、韻律情報調整部27に供給する。

【0091】韻律情報調整部27は、音声韻律情報生成部25から供給される音声韻律情報と、音楽韻律情報生成部26から供給される音楽韻律情報を受信すると、ステップS3において、その音声韻律情報と音楽韻律情報とを調整することで、最終韻律情報を生成し、言語解析部21から供給される音声記号とともに、波形合成部28に出力する。

【0092】波形合成部28は、韻律情報調整部27からの最終韻律情報と音声記号を受信し、ステップS4において、その最終韻律情報と音声記号に基づき、合成音データを生成し、D/A変換部30に出力して処理を終了する。

【0093】これにより、合成音データは、D/A変換部30でD/A変換され、スピーカ18に供給されて出力される。

【0094】従って、ロボットにおいては、音楽的要素により韻律が制御され、ロボットの感情を表現するパラ言語情報が付与された合成音出力される。その結果、ユーザに対して、感情の状態を、合成音によって認識させる、エンタテインメント性の高いロボットを提供することが可能となる。

【0095】以上のように、言語情報に基づいて、合成音の韻律を制御する音声韻律情報を生成するとともに、音楽的要素により韻律を制御することによって、合成音

に対してパラ言語情報を付与する音楽韻律情報を生成し、その音声韻律情報と音楽韻律情報を調整して最終韻律情報を生成するようにしたので、パラ言語情報を、和音や拍子等によって、効果的に伝達することのできる合成音を、容易に生成することができる。

【0096】以上、本発明を、エンターテイメント用のロボット（疑似ペットとしてのロボット）に適用した場合について説明したが、本発明は、これに限らず、例えば、システムに感情等の内部状態が導入された対話システムその他に広く適用することが可能である。

【0097】また、本発明は、現実世界のロボットだけでなく、例えば、液晶ディスプレイ等の表示装置に表示される仮想的なロボットにも適用可能である。

【0098】なお、本実施の形態においては、上述した一連の処理を、CPU10Aにプログラムを実行させることにより行うようにしたが、一連の処理は、それ専用のハードウェアによって行うことも可能である。

【0099】ここで、プログラムは、あらかじめメモリ10B（図2）に記憶させておく他、フロッピー（登録商標）ディスク、CD-ROM（Compact Disc Read Only Memory）、MO（Magnetooptical）ディスク、DVD（Digital Versatile Disc）、磁気ディスク、半導体メモリなどのリムーバブル記録媒体に、一時的あるいは永続的に格納（記録）しておくことができる。そして、このようなリムーバブル記録媒体は、いわゆるパッケージソフトウェアとして提供し、ロボット（メモリ10B）にインストールするようにすることができる。

【0100】また、プログラムは、ダウンロードサイトから、デジタル衛星放送用の人工衛星を介して、無線で転送したり、LAN（Local Area Network）、インターネットといったネットワークを介して、有線で転送し、メモリ10Bにインストールすることができる。

【0101】この場合、プログラムがバージョンアップされたとき等に、そのバージョンアップされたプログラムを、メモリ10Bに、容易にインストールすることができる。

【0102】なお、本明細書において、CPU10Aに各種の処理を行わせるためのプログラムを記述する処理ステップは、必ずしもフローチャートとして記載された順序に沿って時系列に処理する必要はなく、並列的あるいは個別に実行される処理（例えば、並列処理あるいはオブジェクトによる処理）も含むものである。

【0103】また、プログラムは、1のCPUにより処理されるものであっても良いし、複数のCPUによって分散処理されるものであっても良い。

【0104】次に、図4の音声合成部55は、専用のハードウェアにより実現することもできるし、ソフトウェアにより実現することもできる。音声合成部55をソフトウェアによって実現する場合には、そのソフトウェアを構成するプログラムが、汎用のコンピュータ等にイン

ストールされる。

【0105】そこで、図10は、音声合成部55を実現するためのプログラムがインストールされるコンピュータの一実施の形態の構成例を示している。

【0106】プログラムは、コンピュータに内蔵されている記録媒体としてのハードディスク105やROM103に予め記録しておくことができる。

【0107】あるいはまた、プログラムは、フロッピーディスク、CD-ROM、MOディスク、DVD、磁気ディスク、半導体メモリなどのリムーバブル記録媒体111に、一時的あるいは永続的に格納（記録）しておくことができる。このようなリムーバブル記録媒体111は、いわゆるパッケージソフトウェアとして提供することができる。

【0108】なお、プログラムは、上述したようなリムーバブル記録媒体111からコンピュータにインストールする他、ダウンロードサイトから、デジタル衛星放送用の人工衛星を介して、コンピュータに無線で転送したり、LAN、インターネットといったネットワークを介して、コンピュータに有線で転送し、コンピュータでは、そのようにして転送されてくるプログラムを、通信部108で受信し、内蔵するハードディスク105にインストールすることができる。

【0109】コンピュータは、CPU（Central Processing Unit）102を内蔵している。CPU102には、バス101を介して、入出力インタフェース110が接続されており、CPU102は、入出力インタフェース110を介して、ユーザによって、キーボードや、マウス、マイク等で構成される入力部107が操作等されることにより指令が入力されると、それにしたがって、ROM（Read Only Memory）103に格納されているプログラムを実行する。あるいは、また、CPU102は、ハードディスク105に格納されているプログラム、衛星若しくはネットワークから転送され、通信部108で受信されてハードディスク105にインストールされたプログラム、またはドライブ109に装着されたリムーバブル記録媒体111から読み出されてハードディスク105にインストールされたプログラムを、RAM（Random Access Memory）104にロードして実行する。これにより、CPU102は、上述したフローチャートにしたがった処理、あるいは上述したブロック図の構成により行われる処理を行う。そして、CPU102は、その処理結果を、必要に応じて、例えば、入出力インタフェース110を介して、LCD（Liquid Crystal Display）やスピーカ等で構成される出力部106から出力、あるいは、通信部108から送信、さらには、ハードディスク105に記録等させる。

【0110】なお、本実施の形態では、行動決定機構部52が生成するテキスト等の言語情報から合成音を生成するようにしたが、本発明は、あらかじめ用意されたテ

キスト等から合成音を生成する場合にも適用可能である。

【0111】また、本実施の形態では、音楽韻律情報を、感情モデル値に基づいて生成するようにしたが、音楽韻律情報の生成は、感情モデル値以外に、例えば、本能や成長その他のペットロボットの内部状態に基づいて行うことが可能である。

【0112】さらに、音楽韻律情報は、音声韻律情報と同様の記述方法で記述する他、例えば、MIDI (Musical Instrument Digital Interface) の規格に準拠して記述することも可能である。

【0113】

【発明の効果】以上の如く、本発明の音声合成装置および音声合成方法、並びにプログラムによれば、入力データに基づいて、合成音の韻律を制御する第1の韻律情報が生成されるとともに、音楽的要素により韻律を制御し、合成音に対してパラ言語情報を付与する第2の韻律情報が生成される。そして、第1と第2の韻律情報から、最終的な韻律情報が生成され、その最終的な韻律情報により韻律を制御した合成音が生成される。従って、パラ言語情報を、効果的に伝達することのできる合成音を、容易に生成することが可能となる。

【図面の簡単な説明】

【図1】本発明を適用したロボットの一実施の形態の外観構成例を示す斜視図である。

【図2】ロボットの内部構成例を示すブロック図である。

【図3】コントローラ10の機能的構成例を示すブロック図である。

【図4】音声合成部55の構成例を示すブロック図である。

【図5】音声韻律情報と音楽韻律情報の調整方法を説明する図である。

【図6】音声韻律情報と音楽韻律情報の調整方法を説明する図である。

【図7】音声韻律情報と音楽韻律情報の調整方法を説明する図である。

【図8】音声韻律情報と音楽韻律情報の調整方法を説明する図である。

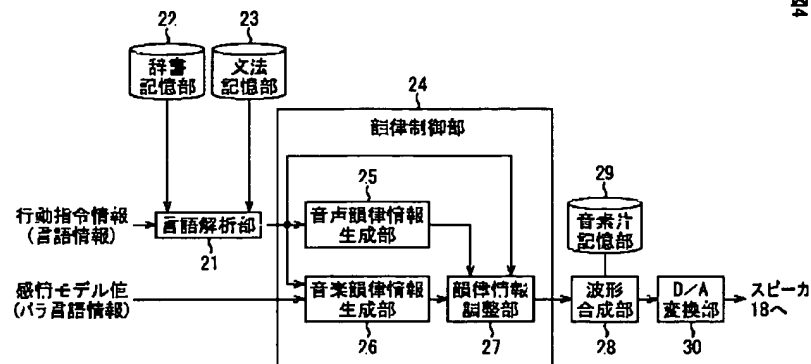
【図9】音声合成部55の処理を説明するフローチャートである。

【図10】本発明を適用したコンピュータの一実施の形態の構成例を示すブロック図である。

【符号の説明】

1 頭部ユニット、 4A 下顎部、 10 コントローラ、 10A CPU、 10B メモリ、 15 マイク、 16 CCDカメラ、 17 タッチセンサ、 18 スピーカ、 21 言語解析部、 22 辞書記憶部、 23 文法記憶部、 24 韻律制御部、 25 音声韻律情報生成部、 26 音楽韻律情報生成部、 27 韻律情報調整部、 28 波形合成部、 29 音素片記憶部、 30 D/A変換部、 50 センサ入力処理部、 50A 音声認識部、 50B 画像認識部、 50C 圧力処理部、 51 モデル記憶部、 52 行動決定機構部、 53 姿勢遷移機構部、 54 制御機構部、 55 音声合成部、 101 バス、 102 CPU、 103 ROM、 104 RAM、 105 ハードディスク、 106 出力部、 107 入力部、 108 通信部、 109 ドライブ、 110 入出力インタフェース、 111 リムーバブル記録媒体

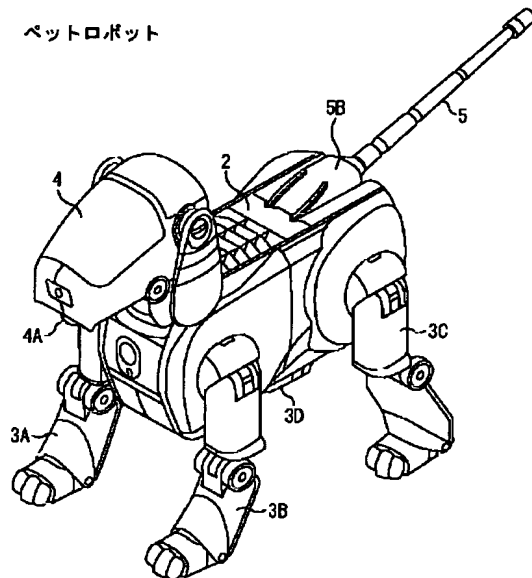
【図4】



【図1】

図1

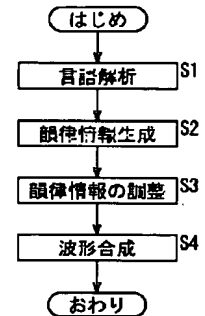
ペットロボット



ペットロボットの外観構成

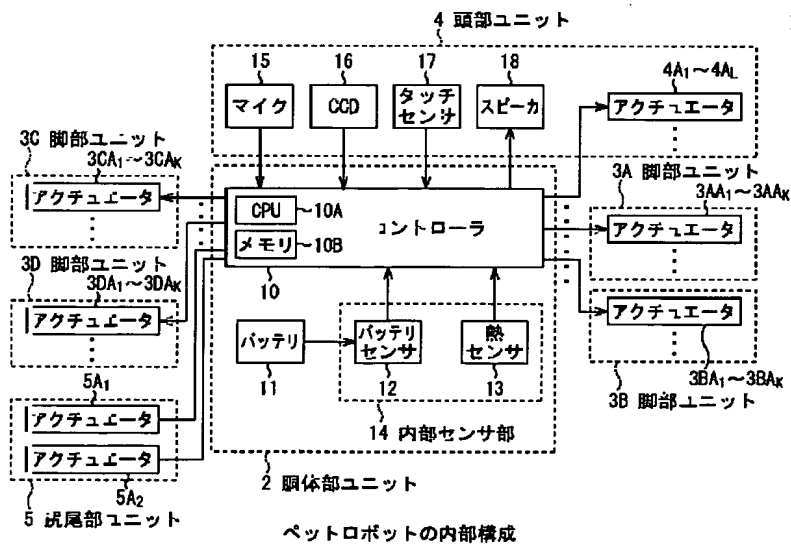
【図9】

図9

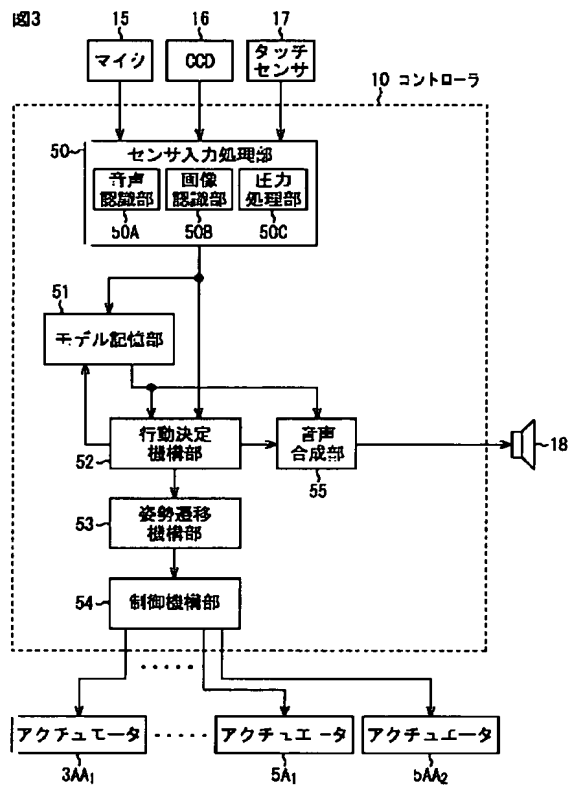


【図2】

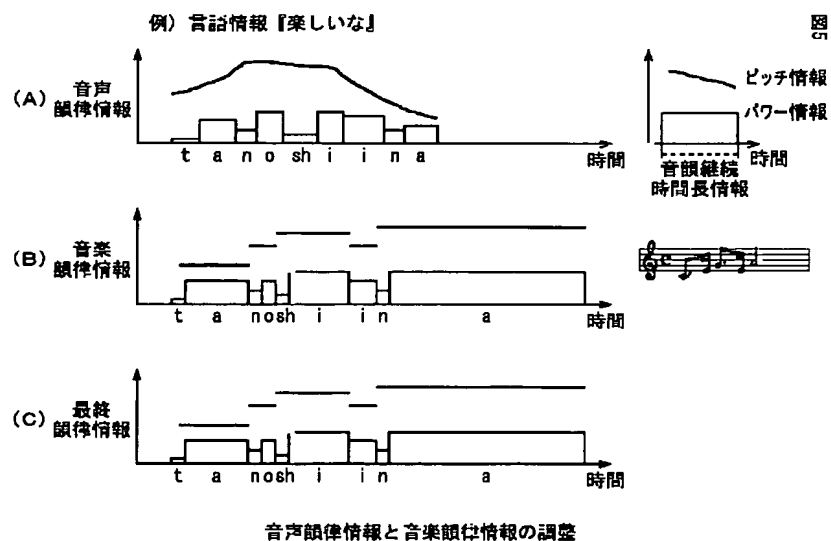
図2



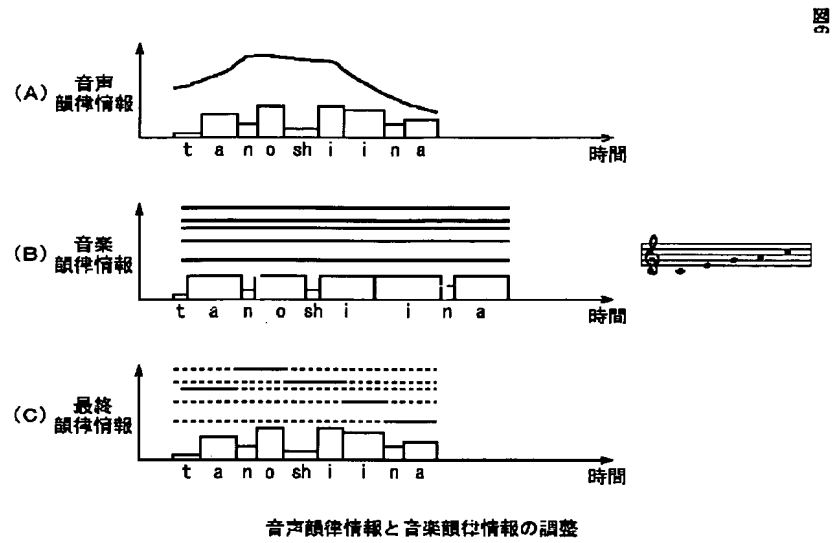
【図3】



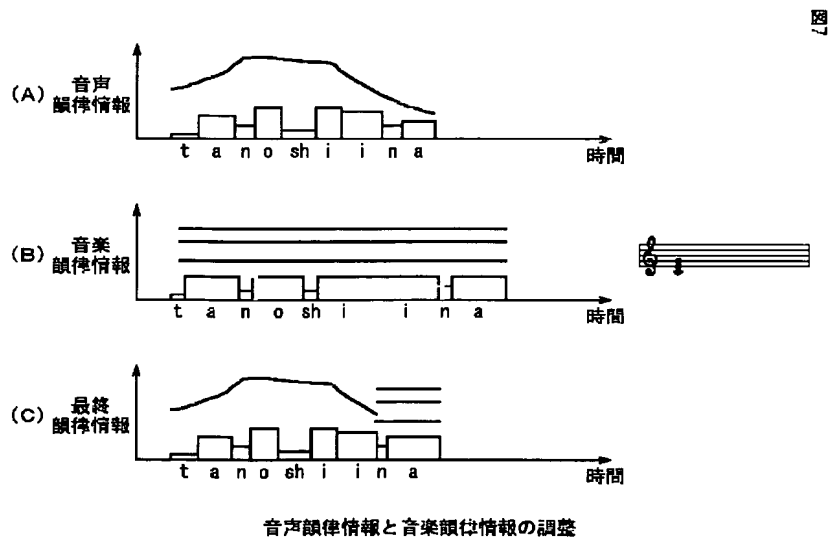
【图5】



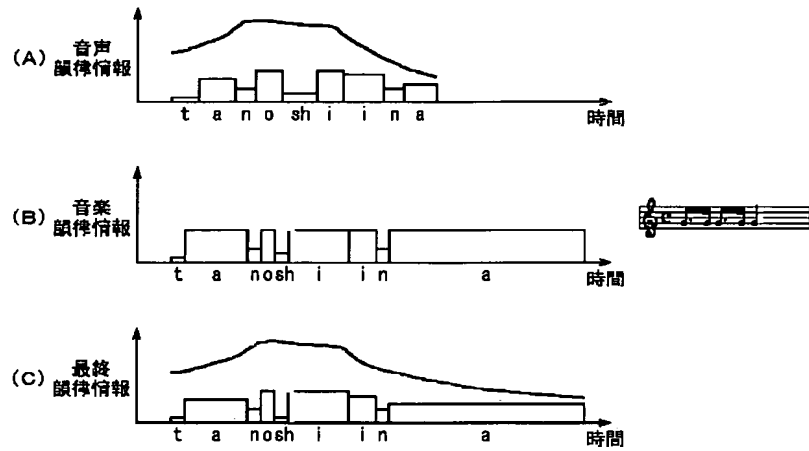
【図6】



【図7】

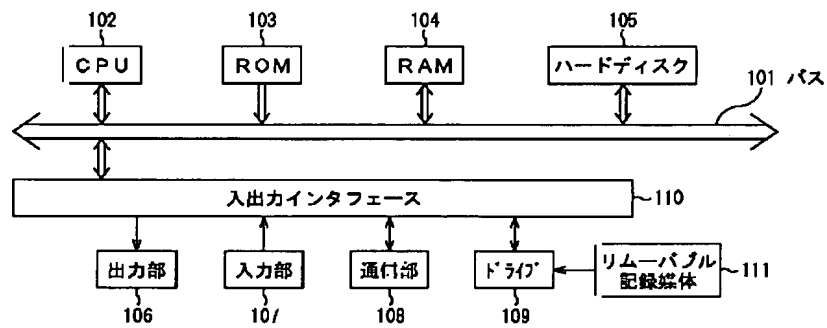


【図8】



音声韻律情報と音楽韻律情報の調整

【図10】



コンピュータ

フロントページの続き

(72)発明者 新田 朋晃  
東京都品川区北品川6丁目7番35号 ソニ  
ー株式会社内

(72)発明者 小林 恵理香  
東京都品川区北品川6丁目7番35号 ソニ  
ー株式会社内

(72)発明者 小林 賢一郎  
東京都品川区北品川6丁目7番35号 ソニ  
ー株式会社内

(72)発明者 赤羽 誠  
東京都品川区北品川6丁目7番35号 ソニ  
ー株式会社内

Fターム(参考) 2C150 CA02 DA05 DA23 DF01 DG22  
DK02 ED08 ED42 ED49 ED52  
ED59 EF11 EF16 EF21 EF23  
EF28 FA01 FA03  
5D015 KK02  
5D045 AA07 AB11